<center>REMARKS</center>

Claims 1-15 and 17-28 are currently pending in the subject application and are presently under consideration. Claims 1, 19, 22, 24 and 26 have been amended as shown on pages 2-5 of the Reply.

Favorable reconsideration of the subject patent application is respectfully requested in view of the comments and amendments herein.

I.   Objection of Claim 19

Claim 19 is objected to because of minor informalities. Withdrawal of this objection is requested for the following reasons. Claim 19 has been amended to cure the minor informalities pointed out by the Examiner. In view of this, the objection is now moot and should be withdrawn.

II.   Rejection of Claim 19 Under 35 U.S.C §112

In the Final Office Action dated March 28 2008, claim 19 stands rejected under 35 U.S.C §112, first paragraph, as failing to comply with the written description requirement. Withdrawal of this rejection is requested for the following reasons. Claim 19 has been amended to cure the minor informalities pointed out by the Examiner. In view of this, the rejection is now moot and should be withdrawn.

III.   Rejection of Claims 24-25 Under 35 U.S.C. §102(e)

In the Final Office Action dated March 28 2008, claims 24-25 stand rejected under 35 U.S.C. §102(e) as being anticipated by Evans (US 2004/0030683). Withdrawal of this rejection is requested for the following reasons. Evans fails to disclose or suggest all features set forth in the subject claims.

> A single prior art reference anticipates a patent claim only if it *expressly or inherently describes each and every limitation* set forth in the patent claim. *Trintec Industries, Inc. v. Top-U.S.A. Corp.,* 295 F.3d 1292, 63 USPQ2d 1597 (Fed. Cir. 2002); *See Verdegaal Bros. v. Union Oil Co. of California,* 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). *The identical invention*

<center>6</center>

> *must be shown in as complete detail as is contained in the ...*
> *claim. Richardson v. Suzuki Motor Co.,* 868 F.2d 1226, 9 USPQ2d
> 1913, 1920 (Fed. Cir. 1989).

Applicants' claimed subject matter provides for a system and method of facilitating incremental web crawls using chunks. Information gathered from a web crawl is indexed and chunked based on similar properties like average time between change and average importance of the retrieved documents. This chunk map then is employed to determine which chunks should be re-crawled. To this end, amended independent claim 24 recites *a data packet transmitted between two or more computer components that facilitates document re-crawl, the data packet comprising: a chunk header that includes metadata associated with the data packet, the metadata shared by all the items in the chunk; an offset section that provides offset information associated with document files; and the document files that include content found on the Internet, wherein the average of the at least one of the properties of all the document files determines if the document should be re-crawled.* Evans is silent regarding such novel features recited by the subject claims.

Evans relates to a system and process for mediated crawling. At the cited portions, Evans discloses a system that searches web sites for target content. The system retrieves media files and data related to media files, on a computer network via a search system utilizing metadata. Auxiliary information associated with each of the websites, stored in a database is utilized to determine a recrawl. When a media file is transmitted, the metadata of that file is transmitted along with it. In contrast, the claimed invention allows for information of documents gathered from a web crawl to be indexed and chunked based on similar properties like average time between change and average importance of the retrieved documents. Each chunk comprises a header that includes metadata that is transmitted, wherein each chunk comprises document files that have similar properties. This allows for transmitting metadata that reflects the average of the properties of all the documents in that chunk. Depending on the metadata, it is determined if all the documents in that chunk should be recrawled. Thus, Evans is silent regarding *a chunk header that includes metadata associated with the data packet, the metadata shared by all the items in the chunk, wherein a chunk comprises document files that have similar properties, the average of the at least one of the properties of all the document files determines if the document should be re-crawled* as recited by the subject claim.

7

From the foregoing, it is clear that an identical invention as recited in the subject claims is not taught or suggested by Evans. Accordingly, it is requested that this rejection with respect to independent claim 24 (and the claims that depend there from) should be withdrawn.

IV.   Rejection of Claims 1-4, 6-15, 17, 19-23 and 26-28 Under 35 U.S.C. §103(a)

In the Final Office Action dated March 28 2008, claims 1-4, 6-15, 17, 19-23 and 26-28 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Najork, *et al.* (US 6,263,364) in view of Evans. Withdrawal of this rejection is requested for the following reasons. Najork *et al.* and Evans, alone or in combination, fail to teach or suggest all features set forth in the subject claims of applicant's claimed invention.

Applicants' claimed subject matter provides for a system and method of facilitating incremental web crawls using chunks. Information gathered from a web crawl is indexed and chunked based on similar properties like average time between change and average importance of the retrieved documents.   This chunk map then is employed to determine which chunks should be re-crawled. To this end, independent claim 1 recites *a system that facilitates incremental web crawls comprising: an indexer that places items with similar properties into respective chunks; and a chunk map that stores at least some of the properties associated with the respective chunk, wherein the properties are at least one of average time between change or average importance of documents comprising a particular chunk, the chunk map employed to facilitate an incremental web re-crawl, wherein the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of that chunk.* Independent claim 26 recites similar features. Najork *et al.* and Evans are silent regarding such novel features.

Najork *et al* relates to a web crawler that downloads documents from a plurality of host computers and enqueues document addresses, where each queue has documents sharing a common host component. At page 5 of the Final Office Action, the Examiner concedes that Najork *et al.* does not teach storing the properties associated with a respective chunk in a chunk map, the chunk map employed to facilitate an incremental web crawl. The Examiner cites Evans to cure the aforementioned deficiencies of Najork *et al.*

Evans relates to a system and process for mediated crawling. At the cited portions, Evans discloses searching web sites for target content. A site map stores contents of the web site,

8

organized in levels wherein each level comprises links, objects, metadata etc related to common content. In a focused recrawl, the site map is used to identify websites containing the requested content, and the identified websites are recrawled. In contrast, the claimed invention allows for chunking, wherein a set of documents that have similar properties are placed in a chunk. Every document shares the properties of the chunk. Each chunk and its properties are stored in a chunk map and this allows for the chunk to be manipulated as one set. Each document in the chunk does not have to be analyzed, but the whole chunk can be recrawled depending on its properties. Further, at the cited portions, Evans discloses storing auxiliary information pertaining to each of the encountered web sites in a database. This stored information is employed to determine how often to recrawl. Thus, the system stores the properties of each and every document in a database. In contrast, the claimed invention allows for chunking, wherein a set of documents that have similar properties are placed in a chunk. Each chunk and its properties are stored in a chunk map and this allows for the chunk to be manipulated as one set. Each document in the chunk does not have to be analyzed, but the whole chunk can be recrawled depending on its properties. Thus, Evans is silent regarding *a chunk map that stores at least some of the properties associated with the respective chunk, the stored properties are shared by all the items in the respective chunk, and the chunk map employed to facilitate an incremental web re-crawl wherein the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of that chunk* as recited by independent claims 1 and 26.

Independent claim 19 recites *a method of performing document re-crawl comprising: parsing a first chunk for uniform resource locators, wherein a chunk map that stores properties associated with the respective chunk is employed to determine the first chunk*, *wherein the stored properties are shared by all the items in the respective chunk; re-crawling the uniform resource locators; and forming a second chunk separate from the first chunk, based at least in part, upon the re-crawled uniform resource locators*. Najork *et al.* and Evans are silent regarding such novel features. At the cited portions, Najork *et al.* discloses performing a recrawl of a queue and enqueueing any new URL's into the front-end queue or in another queue depending on host identifier of the new URL. In contrast, the claimed invention allows for *forming a second chunk*, if the indexer determines that the document belonging to the new URL does not belong to an existing chunk. Thus Najork *et al.* is silent regarding *forming a second chunk separate from the first chunk, based at least in part, upon the re-crawled uniform resource*

locators as recited by independent claim 19. Evans does not compensate for the aforementioned deficiency of Najork *et al.*

In view of at least the foregoing it is readily apparent that Najork *et al.* and Evans, either alone or in combination do not teach or suggest each and every element set forth in the applicants' subject claims. Accordingly it is requested that this rejection should be withdrawn.

V.   Rejection of Claim 5 Under 35 U.S.C. §103(a)

In the Final Office Action dated March 28 2008, claim 5 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Najork, *et al.* in view of Evans and further in view of Eichstaedt, *et al.* (US 6,182,085). It is respectfully requested that this rejection be withdrawn for at least the following reasons. Claim 5 depends from independent claim 1, and as discussed *supra*, Najork *et al.* and Evans, alone or in combination, do not disclose all features of independent claim 1. Eichstaedt *et al.* relates to large scale information gathering utilizing collaborative team crawling, and does not make up for the aforementioned deficiencies with respect to independent claim 1. Thus, the subject invention as recited in the subject claims is not obvious over the combination of Najork *et al.*, Evans and Eichstaedt *et al.* Accordingly, it is respectfully submitted that this rejection of claim 5 should be withdrawn.

VI.   Rejection of Claim 18 Under 35 U.S.C. §103(a)

In the Final Office Action dated March 28 2008, claim 18 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Najork, *et al.* in view of Evans and further in view of Acharaya, *et al.* (US 2007/0094255). It is respectfully requested that this rejection be withdrawn for at least the following reasons. Claim 18 depends from independent claim 1, and as discussed *supra*, Najork *et al.* and Evans, alone or in combination, do not disclose all features of independent claim 1. Acharaya *et al.* relates to document scoring based on link-based criteria, and does not make up for the aforementioned deficiencies with respect to independent claim 1. Thus, the subject invention as recited in the subject claims is not obvious over the combination of Najork *et al.*, Evans and Acharaya *et al.* Accordingly, it is respectfully submitted that this rejection of claim 5 should be withdrawn.

CONCLUSION

The present application is believed to be in condition for allowance in view of the above comments and amendments. A prompt action to such end is earnestly solicited.

In the event any fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP511US].

Should the Examiner believe a telephone interview would be helpful to expedite favorable prosecution, the Examiner is invited to contact applicants' undersigned representative at the telephone number below.

Respectfully submitted,

AMIN, TUROCY & CALVIN, LLP

/Himanshu S. Amin/
Himanshu S. Amin
Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP
24TH Floor, National City Center
1900 E. 9TH Street
Cleveland, Ohio 44114
Telephone (216) 696-8730
Facsimile (216) 696-8731